Measuring Moral Acceptability in E-deliberation: A Practical Application of Ethics by Participation

ILSE VERDIESEN, VIRGINIA DIGNUM, and JEROEN VAN DEN HOVEN,
Delft University of Technology

Current developments in governance and policy setting are challenging traditional top-down models of decision-making. Whereas, on the one hand, citizens are increasingly demanding and expected to participate directly on governance questions, social networking platforms are, on the other hand, increasingly providing podia for the spread of unfounded, extremist and/or harmful ideas. Participatory deliberation is a form of democratic policy making in which deliberation is central to decision-making using both consensus decision-making and majority rule. However, by definition, it will lead to socially accepted results rather than ensuring the moral acceptability of the result. In fact, participation *per se* offers no guidance regarding the ethics of the decisions taken, nor does it provide means to evaluate alternatives in terms of their moral "quality."

This article proposes an open participatory model, Massive Open Online Deliberation (MOOD), that can be used to solve some of the current policy authority deficits. MOOD taps on individual understanding and opinions by harnessing open, participatory, crowd-sourced, and wiki-like methodologies, effectively producing collective judgements regarding complex political and social issues in real time. MOOD offers the opportunity for people to develop and draft collective judgements on complex issues and crises in real time. MOOD is based on the concept of *Ethics by Participation*, a formalized and guided process of moral deliberation that extends deliberative democracy platforms to identify morally acceptable outcomes and enhance critical thinking and reflection among participants.

CCS Concepts: • **Computing methodologies** → *Cognitive science*; • **Human-centered computing** → *Collaborative and social computing systems and tools*; • **Computing methodologies** → **Artificial intelligence**;

Additional Key Words and Phrases: Online deliberation, participatory systems, ethics by participation

ACM Reference format:
Ilse Verdiesen, Virginia Dignum, and Jeroen Van Den Hoven. 2018. Measuring Moral Acceptability in E-deliberation: A Practical Application of Ethics by Participation. *ACM Trans. Internet Technol.* 18, 4, Article 43 (April 2018), 20 pages.
https://doi.org/10.1145/3183324

1 INTRODUCTION

Over the past two decades, the Internet has changed society in unprecedented ways, enabling both new means for collective action and extremes of citizen autonomy and the sharing of radical opinion. It has been the hope of politicians and policy makers that the Internet would fuel and enforce citizen participation in the democratic process. The democratic legitimacy of a political

Authors' addresses: I. Verdiesen, V. Dignum, and J. van den Hoven, Delft University of Technology, Jaffalaan 5, 2628 BX Delft.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 ACM 1533-5399/2018/04-ART43 $15.00
https://doi.org/10.1145/3183324

system and its policies depends on the interaction between citizens and political representatives. Important requirements are a stable communication between policy makers and citizens [3], and the opportunity for citizens to freely discuss and voice their opinions, for which they need to acquire what Dahl has called an "enlightened understanding" of public matters [14]. What we are currently seeing is that the capability to level majority sentiment, traditionally a function of democratic institutions, is eroding under the possibilities of the Internet.

Increasingly, the Internet has rendered the diversity of citizens' views more salient and has proven a powerful medium for discontented citizens to put pressure on the democratic institutions and force changes in policies.[1] People who entertain extremist opinions are now able to connect with like-minded persons and to voice their opinions without any regulation or control. New forms of digitally moderated democratic processes are needed that are able to "filter" or "moderate" extreme opinions, without curbing the right of free speech. This requires methods that enable each participant to evaluate the most relevant opinions for them self, based on both majority and minority voices. In the current democratic system, the views of the minority are suppressed by the victory of the majority. This raises the question if democracy and ethics are compatible and how we could design a system that renders justice to the minority views in a democracy. To be practically adequate as source for deciding policy and political and governance issues, all forms of ethics that are congruent with law need to be designed and implemented in socio-technical systems in the 21st century.

In parallel, current developments in governance are challenging traditional top-down models of decision-making. In fact, authority deficits and legitimacy crises indicate the failure of traditional models of decision-making. Here, participatory processes that promote collaboration among multiple stakeholders are increasingly seen as *the* required alternative to governance fundamental for the success of resource management, (local) government, and international relations. However, the combination of collective inaction, on the one hand, with the increased "politics by social media," on the other hand, can lead to unexpected results and unbalanced participation. Where in general the mild majority keeps silent due to inaction or a sense of uselessness, a few extremist but well-placed one-liners can fuel large crowds.

This can be seen as a sharp increase of feelings of "policy dissonance," i.e., the discrepancy between the values of a community and the perception of those values held by the makers of policies that affect that community. The concept of dissonance was first developed in the social sciences when social psychologist Leon Festinger (1962) forged the concept of *cognitive dissonance* to mean an individual's perception of an incompatibility between conflicting emotions, beliefs, attitudes, or behaviours [20]. Policy dissonance combined with the lack of harmony between what policy makers say or publish and what they do in terms of finance, regulation, or administration is leading to the widening of the gap between politicians and citizens.

The above considerations call for open, distributed, and collaborative models that facilitate participation; encourage the voice of minorities without leading to excesses of unfounded opinions; and, moreover, raise shared awareness of ethical, legal, and social consequences of proposed actions.

Massive Open Online Deliberation (MOOD) environments, as proposed in this article, are structured electronic platforms that tap our individual understanding and opinions by harnessing open, participatory, crowd-sourced, and wiki-like methodologies, effectively producing collective judgements regarding complex political and social issues in real time. As such, MOOD are open-participatory models that can be used to solve some of the current policy authority deficits. The objective is to establish a body capable of making morally authoritative judgements on complex

[1] cf. e.g., https://www.weforum.org/agenda/2016/08/the-biggest-threat-to-democracy-your-social-media-feed/.

issues and crises in real time, based on participatory collaboration and wisdom of the crowd. Wisdom of the crowd refers to the collective opinion of a group of individuals rather than that of a single expert [23]. Relying on wisdom of the crowd has some known drawbacks, such as power concentration in voting behaviour [33], trolling [4], and coordination costs [31], which should be addressed from the start of the design of a participatory platform. There are no intrinsic limits to the issues the initiative could address. Guiding will be several fundamental principles on open participation and formalized deliberation, on which the basic architecture of the platform will be developed.

The article is organised as follows. Section 2 introduces related work, organised into a Learning, Design and hybrid approach to Ethics. Section 3 introduces the concept of Ethics by Participation, used as basis for the participatory platform MOOD, which is presented in Section 4. In Section 5 we describe the results of Citizen Participation Project in Rotterdam. Concluding the article, we discuss our proposal and propose directions for future work in Section 6.

2 RELATED WORK

Making morally authoritative judgements on complex issues are predicated on ethical theories that are increasingly described in Information Technology literature. The Ethics and Autonomous Agents project (eThicAa[2]) distinguishes two theoretical approaches for the logical formalization of ethical models of autonomous agents [5]. One is based on empirical principles, which uses observations of for example common uses and traditions, to build an ethical model. In this light, machine-learning techniques are used by philosophers to learn ethical reasoning [5]. This bottom-up approach is referred to as *Ethics by Learning*. The other theoretical approach is based on deontological ethics that describes what ought to be done to define rules, obligations, and rights. In this view, morality is based on a set of norms and laws that an agent needs to obey. This top-down approach is also referred to as *Ethics by Design*. Both approaches can also be combined to formalize ethical models of autonomous agents in support of a careful moral reflection that is considered essential for ethical decision-making [37]. This hybrid approach is a third approach, additional to the two posed by the eThicAa project. These three approaches are used to classify the articles described in this section. These articles all apply ethical theory to autonomous agents and were chosen because they represent the state-of-the-art literature on this topic. In Section 3, we will contrast these approaches to the one taken in our work: *Ethics by Participation*.

2.1 Ethics by Learning

Malle [34] proposes a framework combining the (up to now) separate fields of robot ethics, in which ethical questions about the design, deployment, and treatment of robots by humans are addressed, and machine morality, which is concerned with questions about the moral capacities of a robot and how these should be computationally implemented. To build a morally competent robot, we need to equip them with a mechanism that allows for *"…constant learning and improvement…"* Malle [34, p. 11]. His thesis is that robots need to learn norms and morality, like little children do, to become morally competent.

Cointe, Bonnet, and Boissier [13] drafted a model in which an agent can judge the ethical aspects of his own behaviour and that of other agents in a multi-agent system. The model describes an Ethical Judgement Process (EJP) that allows agents to evaluate the behaviour of other agents. For this, they need to be able to assess the appropriateness of agents' behaviour with respect to moral convictions and ethical principles. This implies that an agent needs to be aware of other agents' ethical behaviour and adjust its own behaviour according to this.

[2] http://ethicaa.org/.

Davis [15] criticizes the book SuperIntelligence in which Bostrom [7] claims that when an AI system is built with the same intelligence as humans, it will only be a matter of time (years or possibly minutes) before that AI will become immensely more intelligent than humans. As a consequence, it will try to achieve total dominance of our society. Davis proposes that an AI should study a specified collection of dead people (Bostrom assumes that the AI will try to manipulate the ethics of living people) whose ethics you admire and then instruct it to *"Don't do anything that these people would have mostly seriously disapproved of."* [15, p. 123]. It is a good example of how AI could learn a basic set of ethical standards as a sort of default setting on which a more extensive set of values can be learned.

2.2 Ethics by Design

Bonnefon et al. [6] describe the ethical decision, being self-protection or utilitarian, an Autonomous Vehicle (AV) has to made when confronted with pedestrians on the road. In six studies the choices that people would prefer that an AV would make are analysed and their view about governmental regulation is taken into account. The authors state that in some cases, when encountering pedestrians on the road, these AVs will have to make ethical and moral decisions to either protect their passengers or the pedestrians. These choices are based on the trolley problem and are very personal and cultural dependent. Bonnefon et al. [6] state that these decisions will have to be programmed into the AV.

Bryson, Kime, and Zürich suggest that AI can help humanity in improving ethical intuitions and self-understanding, which can lead to better-informed decisions on serious ethical dilemmas [11]. The authors state that AI, and computer programs in general, *"are purpose-built artefacts that are designed, commissioned and operated by humans*" [11, p. 1]. People tend to over-identify with AI, which feeds the fear about this technology. Their conclusion is that we need to consider the ethical and social dangers of AI and address these in the design of AI technologies but only just as much as we would with more conventional technology.

Five years later, Bryson takes an even firmer stance in the discussion on design of AI as moral agents. According to Bryson [10], we are unlikely to construct coherent ethics ourselves, let alone that it is ethical to implement moral subjectivity to AI. She very eloquent states [10, p. 1]: *"We are therefore obliged not to build AI we are obliged to."* By making this argument, she claims that we should not design AI as moral agents at all.

2.3 Hybrid Approaches

In their article, Armstrong, Sandberg, and Bostrom [1] consider a specific type of AI. This so-called Oracle AI only answers questions and does not act that would appear more benign than the *"superintelligent 'genies'"* [1, p. 301] most people think about regarding AI. An Oracle AI can be considered a moral agent in two ways. First, that it has the capacity to draw moral conclusions and, second, as an object of moral considerations itself. The authors propose a combination of *"Direct-Programming, Self-Improvement and Evolution"* [1, p. 302] to develop an Oracle AI as a moral agent. This combination of methods suggests a hybrid approach of Ethics by Design and Ethics by Learning.

Gigerenzer [22] describes the nature of moral behaviour in his essay that, according to the author, is the interplay between mind and environment. He states that it is based on pragmatic social heuristics instead of moral rules or maximization principles. He implies that both nature and nurture are important in the shaping of moral behaviour. Extending this notion to the field of AI implies that both the original programming, which could be seen as the equivalent of nature, and the context, that parallels nurture, will lead to moral agents. Both need to be looked at in conjunction and not as an "either or" decision. This implies that not only an *Ethics by Design*

approach, by means of programming, or *Ethics by Learning*, based on context, suffices for ethical decision making but that a combination of both approaches needed.

2.4 Discussion on Related Work

All articles discussed above take a different approach, by Learning, Design, or a combination of these two, to incorporate the logical formalization of ethical models of autonomous agents. The common denominator is that all these approaches are focussed on ethical reasoning by machines and not by people. To support participatory collaboration and employ the wisdom of the crowd there is a need for an approach that supports *people* in ethical reasoning and assists them in making moral judgements on complex issues. In analogy to the approaches described in this section, we propose to call this approach *Ethics by Participation*.

3 ETHICS BY PARTICIPATION

Participatory deliberation is a form of democratic policy making in which deliberation is central to decision-making using both consensus decision-making and majority rule. Ensuring participation it will lead to socially accepted results but not necessarily ensuring the moral acceptability of the result. Participation *per se* offers no guidance regarding the ethics of the decisions taken, nor provide means to evaluate alternatives in terms of their moral "quality."

Active participation on the debate offers the opportunity for people to develop and draft collective judgement on complex issues and crises in real time. Through a formalized and guided process of (moral) deliberation it will enhance critical thinking and reflection among its participants. This process supports a better understanding of the debated issues, and a clear illustration of the diverse perspectives, while contributing to widely supported solutions. Outcome documents could serve as recommendations for and even exert political pressure on existing traditional authorities.

The fact that large minorities are often being ignored presents a serious issue or that majorities less directly affected force a result detrimental for the directly affected minority is not easy to solve (e.g., in the Colombian Peace referendum in 2016, the negative vote of the urban population that had been much less affected by the guerilla was the reason for the result, overturning the positive vote of the rural populations that had suffered the most). However, through open online deliberation, one can find solutions that consider and integrate various views on certain aspects of a topic. This process has several advantages. It enables people to learn about the different aspects of a complex (political) topic and to better understand each other's positions. At the same time they can contribute to the solution from the beginning, which can lead to a higher satisfaction [39]. Analogously, it should diminish the chances that protest movements and extreme solutions will find good breeding grounds. There can still be a majority vote at the end of a deliberation process. But at this point, the solution will include a substantial amount of the ideas and wishes of all involved participants.

However, despite the enormous democratic and intellectual potential (i.e., the wisdom of the crowd) of an online environment and the exciting opportunities that new social media offer to improve traditional institutions and organisations pertaining to collective decision-making in the common interest of mankind, current online deliberation suffers from two main drawbacks. First, social acceptance (e.g., by means of majority) often differs from moral acceptability, essential for sustainable just decision-making. Second, means are needed to draw and maintain the attention of participants and to help them identify those issues they are most interested in or most qualified to contribute to [32].

One of the main challenges in online deliberation is to harness the wisdom of the crowd and guarantee the quality of collective judgements. What is the right thing to do and why? We know that sometimes propositions are *de facto* accepted that are unacceptable by independent (moral

and epistemic) standards and the available evidence and that other propositions are *de facto* not accepted but that are perfectly acceptable from a moral point of view.

Tapping into the wisdom of the crowd and ensuring the validity of Condorcet's jury theorem, while avoid many of the observed problems in online environments, such as social contagion, availability cascades, filter bubbles, bias, lack of accountability, anonymity, framing, and a number of problems associated with joint action and collective (in)action, MOODs are underpinned by formal structures and reasoning engines to guide sound processes of collective deliberation and reasoning. Based on the practical implementation of deliberative democracy platforms, Fishkin has five characteristics essential for legitimate deliberation [21]:

- **Information:** Accurate and relevant data is made available to all participants.
- **Substantive balance:** Different positions are compared based on their supporting evidence.
- **Diversity:** All major positions relevant to the matter at hand and held by the public are considered.
- **Conscientiousness:** Participants sincerely weigh all arguments.
- **Equal consideration:** Views are weighed based on evidence, not on who is advocating a particular view.

The MOOD platform described in this article is a practical application of *Ethics by Participation* in which these five principles are applied.

4 MOOD

MOOD is based on "collective intelligence," that is, on bringing the knowledge and ideas of many minds together. This combination of ideas and support for constructive discussion has shown to deliver the best results in most challenges [42]. It aims to support free and constructive discussion and integration of several perspectives and thereby benefit several interest groups, not just the incumbent or 51% majority. This is in fact the essence of "digital democracy." The main idea at the basis of the MOOD platform is value-sensitive deliberation. In particular, it embeds the concepts of social acceptance and moral acceptability in the deliberation process. The difference between social acceptance and moral acceptability is that social acceptance is an empirical fact, whereas moral acceptability is an ethical judgement [40].

MOOD aims to structure debates so as to support and enable consensus and to present their outcomes to be used by policy makers to make better decisions. Numerous recent examples show that public debates can be marginally ethical, as they occasionally contain discriminating content. They consist of statements that can be accepted, or not, by a majority of the crowd. For the judgement on the policies, the differentiation between social acceptance and moral acceptability is essential. An example of a policy debate showing the difference between social acceptance and moral acceptability is a debate discussing alternatives to ban certain type of polluting vehicles, such as trucks or old diesel cars, in the city centre to reduce the level of fine dust. The proposal of banning these types of vehicles can be socially acceptable for local residents whose health is at risk but unacceptable for down-town business owners who might lose income. Yet all stakeholders, such as citizens, business owners, customers and suppliers, local government, and environmental defense groups, should reflect on the moral acceptability of banning polluting vehicles from city centres. The moral acceptability, which concerns fairness of decisions but also of the distributions of costs and benefits, the potential future harm to people and environment, risks and control mechanisms, potential oppression, and authority, should be taken into account in this debate. By comparing the social acceptance and moral acceptability of the alternatives discussed in the debate, the policy makers can make tradeoffs in selecting an alternative and, by this, not only look at social acceptance of a proposal but also make an ethical judgement. The level of moral acceptability score, either high or

low, does not imply that an alternative should be selected or not, it merely provides insight into the ethical justness of the alternative. MOOD facilitates this type of debates that should take the views of both the majority as well as the minority into account and strives to be "ethically just," which we define as *stable and sustainable outcomes that are widely accepted.*

We define deliberation as the critical examination of a certain issue where the examination is based on the weighting of pro and con arguments for that issue. A deliberation process allows multiple participants to receive and exchange information, to critically examine this information, and to form a collective judgement (based on the provided information) about a certain issue, which determines the decision-making on a certain issue [19]. MOOD platforms can structure the deliberation process by providing logic to support reasoning, voting procedures, and reputation mechanisms. They can be used by decision makers and citizens to receive the opinions and information from debate participants on certain topics.

To structure a deliberation process we have identified a list of requirements that MOOD platforms should support:

(1) **Demonstration:** enable participation or withdrawing possibilities for participants. Following Reference [27], participants in debates have two alternative ways of reacting to what they perceive as deterioration of the debate or relationship: *exit*, i.e., withdraw as participant, and *voice*, i.e., exert influence for change "from within." As such, MOODs must both enable easy and non-judgemental exit and promote independent and uncriticised voicing of opinions, including those challenging common practice. MOOD platforms should be a forum to talk about values and articulation of reasons.
(2) **Reputation:** promote reputation building of individuals within the community as a reliable signal of the right kind and morally relevant qualities. This reputation should be based on the quality of one's reasons and supporting opinions, judgements, and views and their elaborateness, granularity, relevance, strength, and quality.
(3) **Awareness:** promote awareness of the context of the discussion. This includes both the awareness that something unethical might be going on (i.e., the acceptability of arguments and opinions) but also provide transparency about social relations of participants (i.e., role models, power relations, or authority positions). This feature might be assisted by AI tools; an example of this is Wikipedia, which has a fully autonomous moderation process that ensures that dubious edits are evaluated and, where needed, updated. Introduction of AI is not without problems as the experience with Microsoft's "Tai" chatbot has brought to light. There may also be early-warning tools, for example, to indicate that participants are getting sucked into filter bubbles [17], as there may be indications that participants in their attempts to inform themselves have fallen victim to clickbaits.
(4) **Legitimacy:** Support decision-making based on long-term goals and fundamental shared values rather than on the expediency of the moment and limited self-interest; support participants to move beyond thinking in black-and-white terms about what is right or wrong, and to make a commitment to elevate the discussion to what is best for the community's fundamental values and shared long-term goals. This requires the means for visualisation of "what is best" and suggestions to that effect.
(5) **Perspicuous representation:** this means to transparently render the structure of debates, arguments, logical validity, fact checking but also of the algorithms and models that drive the MOOD environment (meta transparency).

Massive Open Online Deliberation presents the opportunity for people to develop and collectively judge real-time complex issues and crises. Through a formalized and guided process of (moral) deliberation, it will enhance critical thinking and reflection among its participants.

Outcome documents could serve as recommendations for and even exert political pressure on existing traditional authorities. Furthermore, the initiative aims to develop formalized methods of collective (moral) deliberation and the realization of an online reflective equilibrium model, enhancing constant and dynamic feedback on these issues and problems, open participation (crowd sourcing) and potentially even the creation of bottom-up authorities by voluntary association. Last, it could lead to the creation of a new form of communication, discussion, and dispute resolution, which could potentially be used in various settings and on different levels.

4.1 Key Concepts in MOOD

In a MOOD platform several concepts play a key role; the differentiation between facts and values, the concept of moral acceptability and social acceptance, and the e-deliberation process in general. In this section we define these concepts and explain how we implemented these in our prototype.

4.1.1 Facts and Values. A much-debated concept in the world of ethics is the distinction between facts and values. Many philosophers have had their thoughts on how to filter descriptive statements from normative statements. Descriptive, or factual, statements can be used to assert, deny, or communicate about facts. Normative statements, or value judgements, describe how people judge human decisions and conduct and how people value factual matters and circumstances. In our prototype we adhere to this distinction.

The goal of differentiating between facts and values for our system is to have a clear discussion that is based on facts and let participants have a discussion over values that are derived from those facts. Participants will have to provide a fact with the corresponding source, for example, a URL, for every argument they make. To validate the facts and sources provided by participants, we use the methodology of the online encyclopaedia Wikipedia. Wikipedia implemented crowd-sourcing technology, where users (the crowd or editors) have the responsibility of (1) adding content to the encyclopaedia and (2) validating all of the content that is done by panels of experts. Groups of active editors are specialized in certain topics, and if false content on certain pages exists, they will correct this content. In our prototype, this concept is implemented by letting users report on facts that they think are not correct. After reaching a certain amount of so-called report votes, a group of users will be notified to check this fact. This group of users will be randomly selected and they have the responsibility to validate the reported fact or source. If they are not able to judge if a fact is correct or incorrect, then they can inform a group of users who are expert in the field of the source.

To differentiate facts from values, we propose a two step procedure with (1) a randomly selected panel and (2) an expert panel to limit the workload for the expert panel. In other words, the validation of facts in this methodology relies on the wisdom of the crowd. We realize that this methodology might be vulnerable for group-think and strategic behaviour, but we think that Wikipedia proves that the wisdom of the crowd works, if implemented correctly. Another concern is that this procedure will be too slow for a real-time discussion, which is one of the criticisms of the Wikipedia process [30]. As it is hard to keep up in a dynamic debate environment, other means than trying to match this real-time process could be applied. For example, adding a label that the discussion is under investigation as the facts are being validated and inform the participants when a decision on the ruling can be expected. Although this is not a solution for the real-time editing process, it will notify the users on the validity of the facts.

4.1.2 Social Acceptance. Social acceptance can be viewed as a combination of individual feelings, perceived benefits, and risks. It is also a social process in which people are influenced by various types of interactions. Available information and alternative views are important for social acceptance [29]. Research shows that indicators for social acceptance are perceptions of the public, knowledge, and fear [2]. Literature on measuring social acceptance seems to be scarce.

Therefore, we turned to the field of ethics and looked at the Social Choice theory that provides a theoretical framework to reach a collective decision on social welfare. This theory combines individual opinions, preferences, and interests of people. It links welfare economics and voting theory to aggregate preferences and behaviours of individuals. We define social acceptance as the collective decision on the preferences of individuals that is aggregated by means of voting.

To measure social acceptance, we implemented a voting mechanism that is used in Social Choice Theory to determine the social acceptance of the alternatives of the debates. Voting is a popular method to reach a joint decision based on aggregated preferences of individuals. One of the downsides of voting is that the result is sometimes not enough discriminative, as in the Brexit referendum, or even contested, for example, during the 2004 US elections. These downsides can lead to societal unrest and can lead to polarization of the population. A voting mechanism always leads to a clear winner of the majority at the expense of disregarding the preferences of the minority. Many voting systems exist. For example majority and plurality systems where the winner takes all, proportional representation systems where votes are divided over parties and semi-proportional representation systems that include limited and cumulative voting. It is crucial to implement a voting mechanism that incorporates all voters preferences as much as possible before selecting a winner so that the minority feels heard and their preferences taken seriously.

One of the most used voting mechanisms in elections is the Schulze method, which is used by Ubuntu, several Pirate Party political parties, OpenStack, and LiquidFeedback [38]. This preferential voting method satisfies, among others, the criteria of anonymity, the Condorcet criterion, and independence of clones [35]. Voters can list their preferences anonymously that is an important prerequisite for elections. The Condorcet criterion selects a single winner by majority rule in pairwise comparisons over every other candidates. Clone independence is a criterion that prevents certain types of strategic behaviour in the voting process which means that it is impossible to be insincere about a voter's real preferences to secure a more favourable outcome. In the Schulze method, every voter submits an ordered preference list for the candidates presented to the voter. All candidates are compared pairwise and a directed graph with the strongest path is created based on all votes and pairwised comparisons. The output can be determined by looking at which candidate defeated all other candidates, and this one is declared the winner.

Next to the Schulze method, we considered to implement the Ranked Pairs algorithm, because this method is even more robust to strategic behaviour [35], and it satisfies most of the same criteria as the Schulze method. Both are Condorcet methods, but they produce a different order of winners due to the fact that the Schulze algorithm reverses a larger majority than the Ranked Pairs algorithm for the majorities on which the two orders of finish disagree. We found that there is less information available about the Ranked Pairs algorithm than about the Schulze method. Ranked Pairs is also harder to understand, which negatively impacts the transparency of the voting mechanism. Therefore, we chose to implement the Schulze method in our prototype and used the PHP library of Julien Boudry that was available on GitHub [8]. We analysed and tested the implementation of this algorithm with voting example to determine if the open-source algorithm was correct, which it turned out to be.

4.1.3 Moral Acceptability. Morality is concerned with the distinction between right and wrong and contains principles for good and bad behaviour. These principles depend on the political, cultural, and religious context they are defined in Reference [18]. They govern our thoughts, emotions, and behaviour and can be viewed at a personal, interpersonal or collective level [12]. Morality can also be studied on a system level from a more functional approach and can be described as *"Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible."* [25, p. 368]. This systematic approach resulted

in the Moral Foundations Theory, which uses a moral reasoning model based on the principles of *harm, fairness, liberty, loyalty, authority,* and *purity*. We use these principles to define the moral acceptability of the alternatives proposed in the debate process.

In MOOD, moral acceptability is addressed in two alternative ways:

- by means of a Moral Foundations Questionnaire, or
- by facilitation of a discussion on values and value conceptions.

The Moral Foundations questionnaire is useful when large groups are involved, as it can provide an aggregated view of the individual feelings towards the different proposals, but consensus is not *per se* sought. The downside of aggregating views is that individual feelings and views, and thereby the nuance in the debate, can get lost. The value conceptions discussion is more suitable for smaller groups that aim at a consensual view on the solutions. In the following paragraphs, we discuss the two approaches in more detail.

Moral Foundations Questionnaire. One of the approaches used in MOOD to survey the moral acceptability of the alternatives discussed in the deliberation process is based on the Moral Foundations Questionnaire (MFQ) that was developed based on the Moral Foundation Theory (MFT) [25]. The MFT is a pluralistic approach to study morality. It is comprised of five foundations: (1) the Care/harm foundation, (2) the Fairness/cheating foundation, (3) the Loyalty/betrayal foundation, (4) the Authority/subversion foundation, and (4) the Sanctity/degradation foundation. These five foundations have been used to test the MFT in multiple cultures using different samples and research found that the foundations can be used across cultures to detect both cross-cultural as within-cultural differences [25]. The downside of the MFT is that it is limited to these five principles and it does not consider other moral values, for example, on ecology or privacy. However, we believe that the MFT is suitable to investigate the moral judgement in e-deliberation due to its pluralistic and universal nature that allows for application in different cultural groups and contexts. With all studies in social sciences, one has to be cautious of comparing inter-cultural results, because not only the context is different, also the sample of participants is construed of different features, for example the ratio men-women or distribution over age groups.

The foundations of the MFT have been translated in a questionnaire (MFQ) that can be used to measure a broad range of moral concerns. The MFQ consists of two parts, one on moral relevance and the other one is about moral judgement. It contains questions to get insight into the respondent's view on the principles of *harm, fairness, liberty, loyalty, authority,* and *purity*. To test the applicability of MFT for the aims of MOOD, we selected the principles of *harm*, *fairness*, and *authority* to implement in our prototype. These are values mentioned by Van de Hoven et al. [28] as considerations of moral importance in their description of the concept of *Privacy by Design* for ICT-driven developments. The MOOD platform can be viewed as an example of an ICT-driven development of deliberation, and therefore we selected the same values to be applied in our prototype. In future work, we will extend the prototype to include all six principles and to use the 15 questions of the first part as an instrument to assess the moral acceptability of the proposed alternatives in the debates.

We ran a test with fellow students to check the understandability of the questions. It turned out that the questions in their original form were hard to understand by the testers and did not make sense to them when applied to the alternatives. Therefore, we decided to adjust the MFQ questions slightly to make them more applicable to our design of the debate process and understandable for the user. An example of this modification is the rephrasing the statement in Table 1: *Whether or not some people were treated differently than others* into the question: *Do you think that as a result of the alternative above: Someone is treated differently from others?* We realize that this impacts the

Table 1. Overview Transcription Original MFQ Questions into MOOD Questions

Original MFQ question	**Transcribed MOOD question**
When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking?	Do you think that as a result of the alternative above:
Harm principle	
1. Whether or not someone suffered emotionally.	1. Someone suffers emotionally?
2. Whether or not someone cared for someone weak or vulnerable.	2. Someone cares for someone weak or vulnerable?
Fairness principle	
3. Whether or not some people were treated differently than others.	3. Someone is treated differently from others?
4. Whether or not someone acted unfairly.	4. Someone acts unfairly?
5. Whether or not someone was denied his or her rights.	5. Someone is denied his or her rights?
Authority principle	
6. Whether or not someone showed a lack of respect for authority.	6. Someone shows a lack of respect for authority?
7. Whether or not an action caused chaos or disorder.	7. Someone takes action that will cause chaos or disorder?

validity of this instrument that means that research is needed to validate the modified questions. Since our prototype is merely a proof of concept, we chose not to test this validity at this moment. In the text below, the MFQ questions for the *harm*, *fairness*, and *authority* principles are applied to an example debate. Building on Figure 3, the debate question is *"What should the town counsel do to improve the air quality in the city center?"* The proposed alternatives are "Electric cars" and "Low Emissions Zone." Debaters are asked to answer the MFQ questions for both alternatives on a five-point Likert scale ranging from "strongly agree" (1) to "strongly disagree" (5). The alternatives are presented at random to the participants for the moral acceptability testing and this presentation is not based on the ranking as result of the voting. In this way, we prevent that the moral acceptability scores are highly dependent on, or impacted by the earlier ranking of the alternatives. This means that not only the alternatives favoured by the majority are tested for moral acceptability, but also the alternatives that represent the minority views. All participants will be able to see the results of the moral acceptability of the alternatives afterwards and this includes the results of the minority alternatives. It might very well be that the results of the moral acceptability will change people's opinions regarding their own moral acceptability score of the alternatives and may even lead to a change of their preferences. During the G1000 citizen participation event (Section 5) we noticed that this was actually the case. In future implementation of the MOOD systems, we would like to study the effect of the overall moral acceptability score on the participant's preferences.

Questions 1 and 2 refer to the *harm* principle, 3, 4, and 5 to the *fairness* principle and questions 6 and 7 represent the *authority* principle. In the current version of MOOD, we have chosen a simple average to present the aggregated results of these questions. These are presented in

Figure 4. Future work is needed to evaluate different aggregation functions, their usability, and understandability from the perspective of the users.

Value Conception Discussion. Given that the aim of the MOOD process is support participants achieve a better understanding of others' perspectives, taking values as the focus of the discussion can ensure that conflicts might be less severe or even prevented [9, 24]. This requires that participants understand what each mean by a certain value or what one's *value conceptions* are. Rawls [36] states that understanding the existing overlap between the different conceptions leads to a generally supported conception of the value at stake [36]. Habermas [26] argues that stakeholders should strive for mutual understanding of the moral points of view by discussing the different perceptions open and freely [26].

MOOD supports this process by enabling participants to formulate and, consequently, discuss the values they associate with the different alternatives being discussed. This value discussion follows an iterative process, roughly based on the Delphi method:

- Participants are asked to formulate which values are relevant for each of the alternatives. This includes both those values that are promoted by the alternative as those that are possibly demoted;
- Participants then describe the reasons behind the values they've listed, and discuss how they perceive those values;
- As these perceptions can be very different, the important aspect here is to allow for understanding and acceptance of each others perspectives;
- After this discussion, participants are asked to rank the alternatives a second time
- Differences in ranking are then discussed.

4.2 Practical Application of MOOD

In our prototype,[3] we implemented an actual deliberation process that consists of four stages: (1) proposal and initiation of a debate, (2) the actual debate in which user can cast votes to support an alternative, (3) the selection of alternatives via preference voting and measuring the moral acceptability of the alternatives, and (4) reporting of the results. These stages are depicted in Figure 1, which are translated to the application in the overview of the debate page in Figure 2.

In stage 1, a user can initiate a debate by submitting a proposal to the MOOD platform. This proposal needs to be described in a generic way and should preferably be posed as an open question. The initiator has four weeks to raise support for the debate and to reach a voting threshold. The threshold procedure resembles the procedure for citizen initiatives in The Netherlands [16]. After reaching the voting threshold the proposal enters stage 2 of the debate. Once the threshold is reached, an initiator cannot withdraw his or her proposed debate, because this would mean that all aspects of a certain debate, like arguments, sources, and facts, will be deleted and, in our opinion, valuable information will be lost.

In stage 2, the actual debate is held. Discussants can react to a debate by submitting alternatives that consist of pro and con arguments (Figure 3). It is also possible for users to add pro or con arguments to an existing alternative. Arguments need to be substantiated by facts and sources to reference these facts to differentiate them from values. Although not built into our prototype yet, these facts will be validated by means of crowd-sourcing. The facts can be contested by other users and if a certain threshold is reached, then the administrator will review the fact. If the fact is not valid, then it will be marked in the database as rejected and will not be visible to the users. In a future version of the MOOD platform an expert panel will take over this task from

[3] https://mood.tbm.tudelft.nl/.

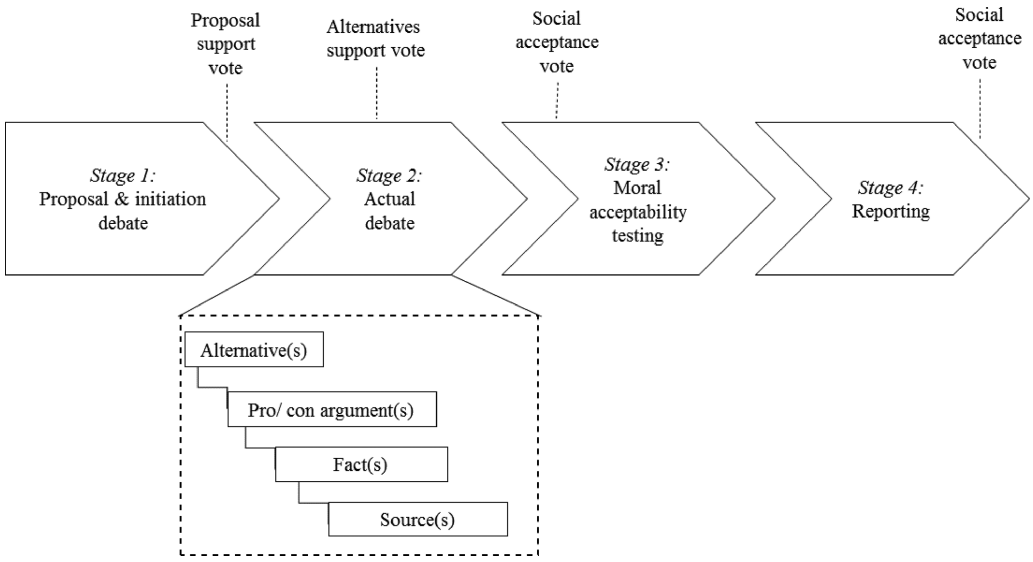

Fig. 1. MOOD deliberation process.

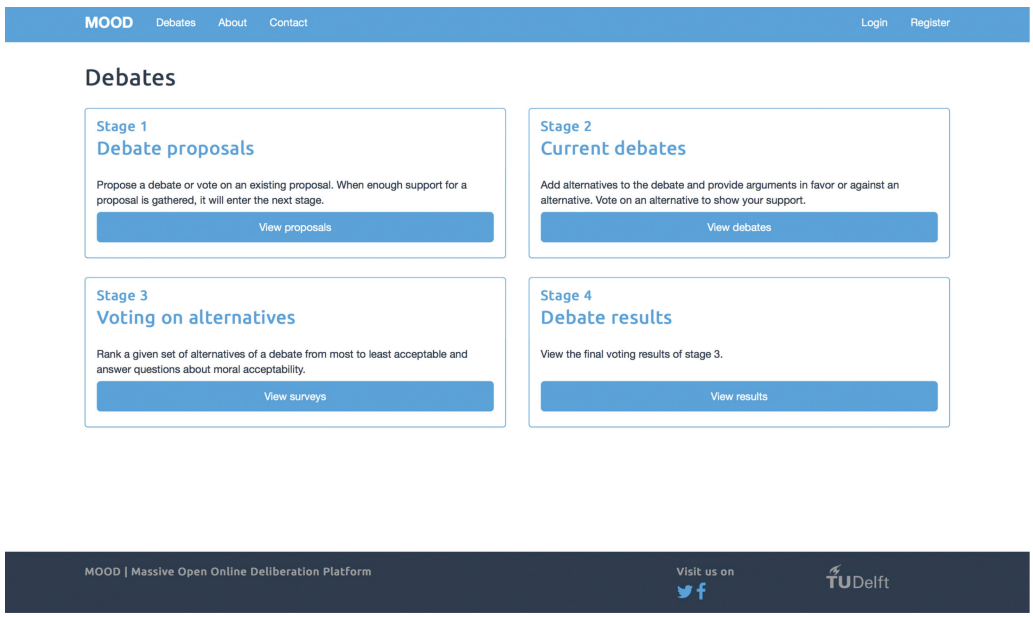
MOOD
Debates
About
Contact
Login
Register
Debates
Stage 1
Debate proposals
Propose a debate or vote on an existing proposal. When enough support for a proposal is gathered, it will enter the next stage.
View proposals
Stage 2
Current debates
Add alternatives to the debate and provide arguments in favor or against an alternative. Vote on an alternative to show your support.
View debates
Stage 3
Voting on alternatives
Rank a given set of alternatives of a debate from most to least acceptable and answer questions about moral acceptability.
View surveys
Stage 4
Debate results
View the final voting results of stage 3.
View results
MOOD | Massive Open Online Deliberation Platform
Visit us on
TUDelft

Fig. 2. Screenshot debate page.

the administrator to provide a more independent judgement of a contested fact. A debate will have a pre-set duration that is set by the initiator. In this stage, all users can vote to support an alternative. The five top alternatives will be selected and the debate will enter the next phase.

In stage 3 of the debate, a voter can list his or her preferences of alternatives. The preferences are calculated by the Schulz voting mechanism. By this, the social acceptance of the alternatives in a debate is measured. After the voting, a list of alternatives is created ranking the alternatives

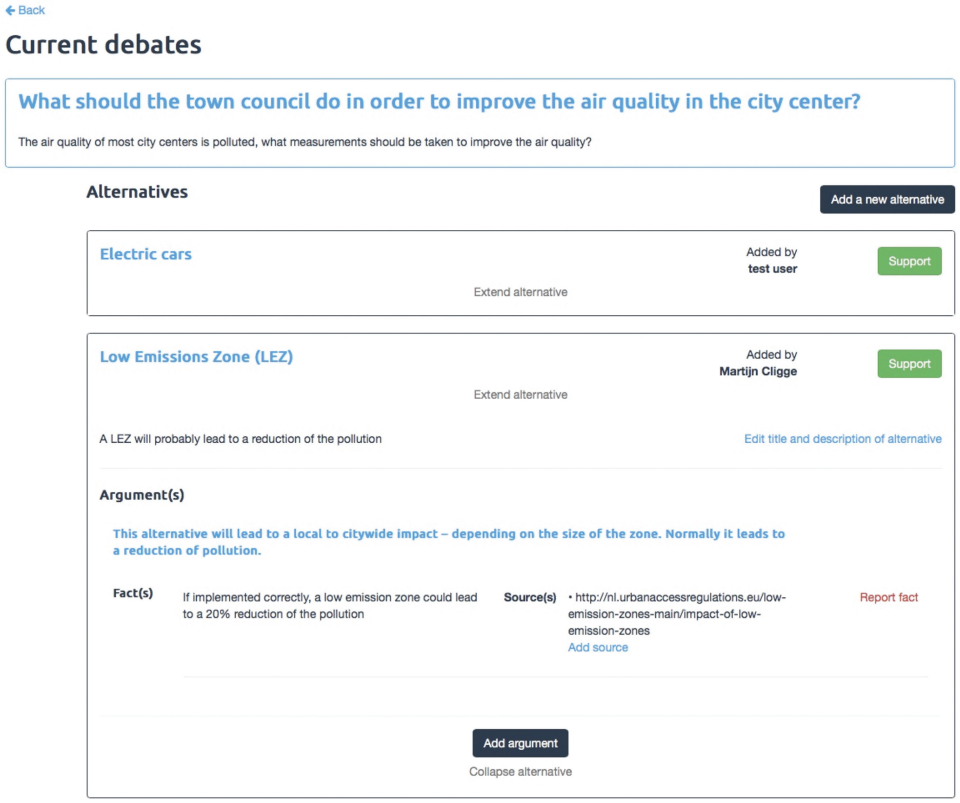
Back
Current debates
What should the town council do in order to improve the air quality in the city center?
The air quality of most city centers is polluted, what measurements should be taken to improve the air quality?
Alternatives
Add a new alternative
Electric cars
Added by
test user
Support
Extend alternative
Low Emissions Zone (LEZ)
Added by
Martijn Cligge
Support
Extend alternative
A LEZ will probably lead to a reduction of the pollution
Edit title and description of alternative
Argument(s)
This alternative will lead to a local to citywide impact – depending on the size of the zone. Normally it leads to a reduction of pollution.
Fact(s)
If implemented correctly, a low emission zone could lead to a 20% reduction of the pollution
Source(s)
• http://nl.urbanaccessregulations.eu/low-emission-zones-main/impact-of-low-emission-zones
Add source
Report fact
Add argument
Collapse alternative

Fig. 3. Screenshot alternatives page.

that received the most votes. Next, the moral acceptability of the alternatives is either surveyed using the MFQ, or identified using the value perception discussion. This step aims at gain insight in the moral acceptability of the different alternatives in a debate.

In stage 4, the final stage the social acceptance and moral acceptability results of the debate will be presented (Figure 4 depicts the results for the MFQ case). The results will be available to all users, which will enhance the transparency of the debate.

4.3 Ethical Participation in MOOD

The MOOD prototype described above adheres to the five characteristics that according to Fishkin [21] are essential for legitimate deliberation. The platform *informs* all participants and makes all deliberation arguments and data available for people who have an account, but also non-registered users can view the outcomes of the debates. By means of the URL, field evidence is required to support positions, which means that the platform promotes *substantive balance* between opinions and facts. *Diversity* is stressed in that debaters can provide both pro as con arguments. By presenting all users the same information before the social acceptance vote the participants can *conscientiously* weigh all arguments. The Condorcet voting method protects the minority viewpoints and therefore the MOOD prototype abides by the *equal consideration* characteristic.

The structure of the phases in the deliberation process in a MOOD platform is aimed at enhancing critical thinking and reflection among the participants in a threefold way. First, each participant is presented the score of his or hers individual MFQ. This will mirror to each of the participants

Fig. 4. Screenshot MFQ results page.

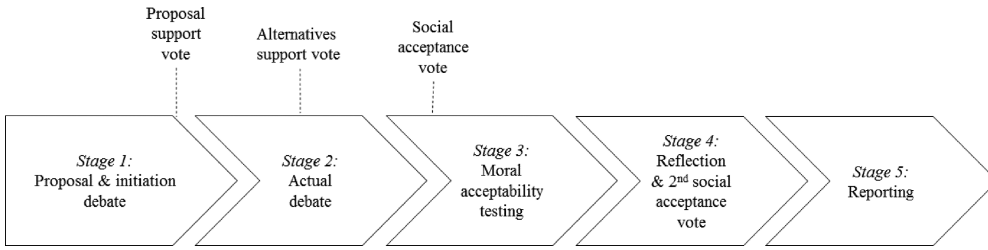
Proposal support vote
Alternatives support vote
Social acceptance vote
Stage 1: Proposal & initiation debate
Stage 2: Actual debate
Stage 3: Moral acceptability testing
Stage 4: Reflection & 2nd social acceptance vote
Stage 5: Reporting

Fig. 5. Updated MOOD deliberation process.

how the preferred alternatives hold against their own moral values. Second, a collective MFQ will aggregate the answers of all participants and the collective MFQ can give an indication of the shared ethical stances of the population that participated in the debate as a whole. Finally, it would be possible to vote again on the alternatives after people have seen the MFQ results. This may lead to other preferences that might arguably more "ethical." In the MOOD prototype described above, the individual MFQ mirror is implemented. The collective and second social acceptance voting is not implemented yet in the prototype but is a feature that needs to be added in a next version. Based on the *Ethics by Participation* approach, the initial MOOD deliberation process would be updated with a second social acceptance vote round as shown in Figure 5.

5 CITIZEN PARTICIPATION IN ROTTERDAM

In July 2017, the MOOD deliberation process has been used during the G1000 citizen participation event organised by the city of Rotterdam.[4] The platform has also been applied to several debates

[4] http://www.g1000rotterdam.nl/.

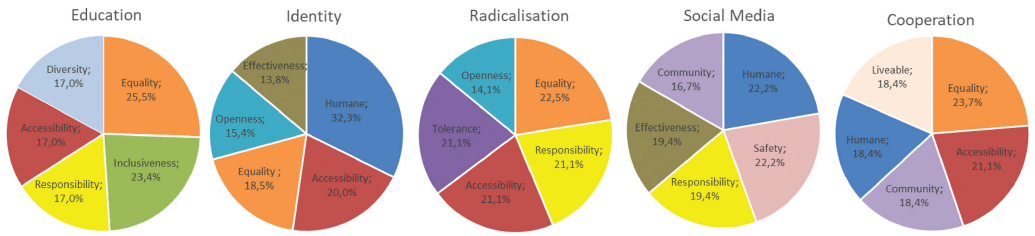

Fig. 6. G1000 Rotterdam: Main values identified per topic.

Table 2. Change of Ranking for the Issue "How to Make Educators Aware of What Skills Children Need to be Able to Fully Participate in Society?" in the Topic Education

Alternative	1st solutions ranking	2nd solutions ranking
A	2	1
B	3	3
C	1	2
D	4	4

regarding water management projects in The Netherlands. However, these were small-scale events that can best be seen as proof of concept rather than full-fledged application. In the case of the G1000 event, 1500 citizens of Rotterdam were invited to use the MOOD deliberation process to discuss their values and generate solutions in the topics of Education, Cooperation, Social Media, Radicalisation, and Identity. In this event, a combination of face-to-face discussion and online structuring of the debate, following the four phases described in Section 4.2 was used. Moreover, we used the Value Conception Discussion approach to the value identification phase. Participants worked in groups and in total identified 95 issues they considered worthy of discussion, involving 616 active participants. Issues were distributed between the five topics as follows:

- education: 12 issues
- identity: 16 issues
- radicalisation: 14 issues
- social media: 12 issues
- cooperation: 11 issues

For each issue, participants were asked to identify four possible alternatives (including the alternative *"do nothing"*). Following the Value Conception Discussion described in Section 4.1.3, after listing possible alternative answers to the issues, all participants were asked to rank these alternatives according to their preferences. This was followed by the discussion on the values that participants identified as being pertinent to the issue. Following this discussion, participants were asked to rank again the alternatives.

Overall, the top 10 values identified in the discussions were *Equality, Accessibility, Humanity, Tolerance, Responsibility, Effectiveness, Inclusiveness, Safety, Diversity,* and *Openness*. Figure 6 shows the distribution of the top 5 values for the five topics of the G1000 debate (Education, Identity, Radicalisation, Social Media, and Cooperation).

For example, in the topic Education, one of the discussed issues was *"How do we make educators aware of what skills children need to be able to fully participate in society?"* In this case, the discussion identified the following four alternative solutions. Table 2 shows the changes in rankings

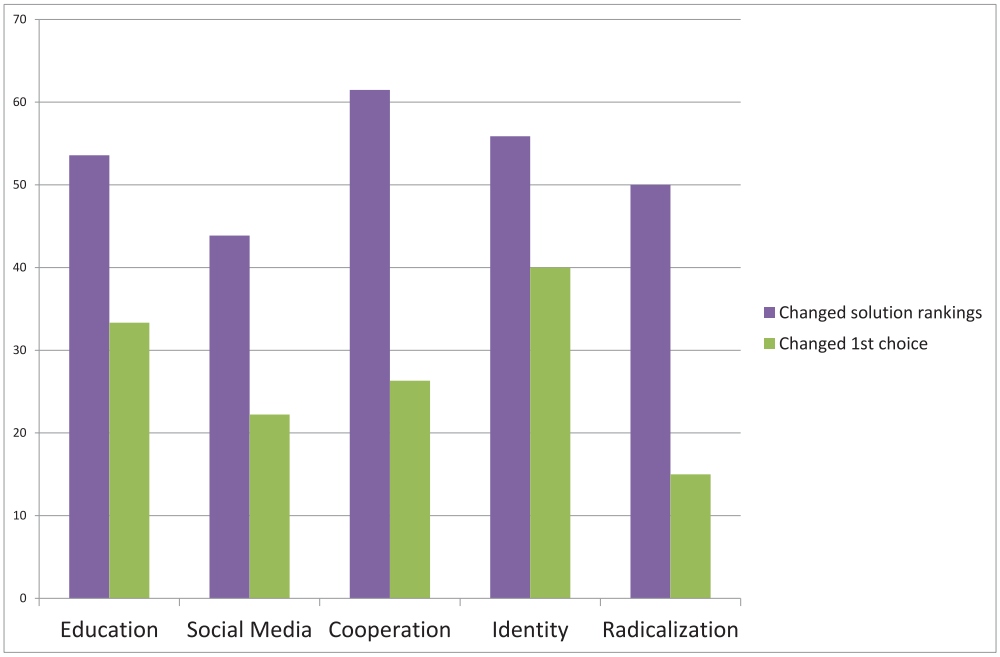

Fig. 7. G1000 Rotterdam: Changes due to value discussion.

of the solutions identified for this issue, using Condorcet aggregation, for all participants in this discussion. In most cases, as individual participants change their rankings this has an effect on the final preferred solution.

A. Organise drama lessons at school about social themes with the participation of parents.
B. Provide parents clear information about methods and objectives of school education to manage expectation.
C. Invest on extra dialogue between school and parents.
D. Do nothing.

Figure 7 gives an overview of the preliminary results, indicating per topic the percentage of changed ranking overall and of participants who changed their most preferred option. For example, in the topic Education, 53% of participants changed their rankings (across all the Education issues discussed), and 33% changed their most preferred alternative (for the issue they discussed). In 15 of the 65 discussed issues, the first choice was altered after the values discussion. Moreover, in total, 50% of the participants altered their own rankings in some way after the discussion. Most changes where in the discussion of issues in the "Radicalisation" topic (54% of participants modified their rankings after the value discussion).

Finally, we asked all participants to fill out a questionnaire reporting on their own evaluation of the discussion. In total, 303 participants filled out the questionnaire, overwhelmingly reporting a better understanding of the position and feelings of the other participants as depicted in Table 3 (split into participants that have changed their own preferences between the first and second rankings and those who remained equal).

Even though results are preliminary, and only based on one large case study, they show the potential of the approach in terms of individual reflection and the effect of attempting to understand

Table 3. Changed Understanding as Reported by Participants in the Exit Questionnaire

	Changed	Not changed
Less understanding	0%	0.5%
Equal understanding	23%	39%
More understanding	77%	60.5%

the values and value conceptions held by others. By focussing the discussion on these values, the results achieved are supported by a wider group.

6 CONCLUSION AND DISCUSSION

In this article, we have presented a novel *Ethics by Participation* approach for participatory deliberation that enables discussion and measures the moral acceptability of complex issues. This approach is aimed at enhancing critical thinking and reflection among debate participants and taps into the intellectual potential of the wisdom of the crowd. As such, more than supporting the election of a particular solutions, the MOOD process aims at induce a better understanding of each others positions and values and thus a better discussion and reflection in the decision by the "crowd." The article provides a description of the overall method and touches on the ongoing evaluation of the MOOD platform.

6.1 Limitations

Given the preliminary nature of the MOOD platform, several limitations can be identified. Although based on often described approaches, *Ethics by Participation* is a new construct and therefore minimally supported by theory. This also applies to the measurement of moral acceptability, on which little research was found. The MFQ that was implemented in the prototype is normally used as a questionnaire to get insight into personal moral views and is not intended to morally judge on complex issues. Moreover, MFQ investigates peoples appreciation of a set of pre-established values. In the implementation of MOOD described in this article, this concerns the values harm, fairness, and authority. However, the approach does not allow us to test other values that a participant may find important, such as, e.g., sustainability or privacy. Another limitation of our approach is the method to differentiate facts from values. In the prototype, attaching a URL was a sufficient substantiation as indication for a fact. However, attaching a URL does not necessarily point to a fact as not all information on the Internet is, by definition, factual. Finally, the proposed deliberation process is currently being used in real-life experiments, but the results need to be studied to validate MOOD's applicability in practice and to study the effect of the drawbacks of the wisdom of the crowd approach, such as trolling and power concentration of voting behaviour, on the deliberation process.

6.2 Further Research

The limitations discussed above call for further research in that the proposed deliberation process needs to be validated in substantive research. One of the future experiments must address the known drawbacks of the wisdom of the crowd to check whether power concentration in voting behaviour occurs and if the design of the moderation is resistant to trolling. In a future implementation, the effect of taking note of the overall moral acceptability score on participant's own preferences should be studied to check whether this will lead to a change in their preferences as the G1000 citizen participation event in Rotterdam seems to indicate. Also, a future implementation should encompass all six MFT values and 15 questions of the first part of the MFQ to meet all

cross-cultural aspects that are incorporated in the Moral Foundations Theory. In further studying the differentiation between facts and values, it could be advisable to turn to the field of investigation, because in criminal investigations, and court cases, all evidence is based on facts. It could be beneficial to study these evidence processes and apply these to the MOOD deliberation process. Last, the *Ethics by Participation* approach needs further evaluation and theoretical grounding as basis for participatory deliberation.

Currently, we use the outcome of the tool as recommendations for policy-makers and as direct decisions. Future developments include formalized methods of collective (moral) deliberation and the realization of an online reflective equilibrium model, enhancing constant and dynamic feedback on complex issues and problems (real-time), open participation (crowd sourcing) and potentially even the creation bottom-up authorities by voluntary association. These results must be evaluated in different contexts and decision areas but can eventually lead to the creation of a new form of communication, discussion, and dispute resolution, which could potentially be used in various different settings and on different levels.

ACKNOWLEDGMENTS

We acknowledge Martijn Cligge, Jan Timmermans, and Lennard Segers, who co-developed the prototype of the MOOD platform and co-authored the article "MOOD: Massive Open Online Deliberation Platform—A practical application" [41]. We also thank Klara Pigmans for the analysis of the G1000 Rotterdam results and Lokaal Rotterdam for its organization.

REFERENCES

[1] Stuart Armstrong, Anders Sandberg, and Nick Bostrom. 2012. Thinking inside the box: Controlling and using an oracle AI. *Minds Mach.* 22, 4 (2012), 299–324.
[2] Getachew Assefa and Björn Frostell. 2007. Social sustainability and social acceptance in technology assessment: A case study of energy technologies. *Technol. Soc.* 29, 1 (2007), 63–78.
[3] Arthur Benz, Ioannis Papadopoulos, and others. 2006. *Governance and Democracy: Comparing National, European and International Experiences*. Routledge.
[4] Jonathan Bishop. 2014. Representations of 'trolls' in mass media communication: A review of media-texts and moral panics relating to 'internet trolling'. *Int. J. Web Based Commun.* 10, 1 (2014), 7–24.
[5] Olivier Boissier, Grégory Bonnet, Jean-Gabriel Ganascia, Catherine Tessier, and Robert Voyer IMT. 2015. A roadmap towards ethical autonomous agents. (2015).
[6] Jean-Franois Bonnefon, Azim Shariff, and Iyad Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352, 6293 (2016), 1573–1576.
[7] Nick Bostrom. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Oxford.
[8] Julien Boudry. Condorcet. Retrieved May 26, 2016 from https://github.com/julien-boudry/Condorcet.
[9] Robert O. Briggs, Gwendolyn L. Kolfschoten, and Gert-Jan de Vreede. 2005. Toward a theoretical model of consensus building. In *Proceedings of the Annual Americas Conference on Information Systems (AMCIS'05)*.
[10] Joanna J. Bryson. Patiency is not a virtue: AI and the design of ethical systems. In *Proceedings of the 2016 Conference on Artificial Intelligence Spring Symposium Series (AAAI'16)*.
[11] Joanna J. Bryson, Philip P. Kime, and C. H. Zúrich. 2011. Just an artifact: Why machines are perceived as moral agents. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* 22, 1(2011), 1641.
[12] Taya R. Cohen and Lily Morse. 2014. Moral character: What it is and what it does. *Res. Org. Behav.* 34 (2014), 43–61.
[13] Nicolas Cointe, Grgory Bonnet, and Olivier Boissier. Ethical judgment of agents' behaviors in multi-agent systems. In *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1106–1114.
[14] Robert Alan Dahl. 1989. *Democracy and Its Critics*. Yale University Press. 111–112.
[15] Ernest Davis. 2015. Ethical guidelines for a superintelligence. *Artif. Intell.* 220 (2015), 121–124.
[16] Tweede Kamer der Staten Generaal n.d.Burgerinitiatief. Retrieved July 7, 2016 from https://www.tweedekamer.nl/kamerleden/commissies/ver z/burgerinitiatieven.
[17] Bozdag E. and J. van den Hoven. 2015. Breaking the filter bubble: Democracy and design. *Ethics and Information Technology* 17, 4 (2015), 249–265.

[18] Naomi Ellemers, Stefano Pagliaro, and Manuela Barreto. 2013. Morality and behavioural regulation in groups: A social identity approach. *Eur. Rev. Soc. Psychol.* 24, 1 (2013), 160–193. DOI: http://dx.doi.org/10.1080/10463283.2013.841490 arXiv:http://dx.doi.org/10.1080/10463283.2013.841490
[19] James D. Fearon. 1998. Deliberation as discussion. *Deliberative Democracy* 44 (1998), 56.
[20] Leon Festinger. 1962. *A Theory of Cognitive Dissonance*. Vol. 2. Stanford University Press.
[21] James S. Fishkin. 2011. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford University Press.
[22] Gerd Gigerenzer. 2010. Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science* 2, 3 (2010), 528–554.
[23] Jim Giles. 2005. Wisdom of the crowd. *Nature* 438, 7066 (2005), 281.
[24] L. Glenna. 2010. Value-laden technocratic management and environmental conflicts: The case of the new york city watershed controversy. *Sci. Technol. Hum. Values* 35, 1 (2010), 81–112.
[25] Jesse Graham, Brian A. Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H. Ditto. 2011. Mapping the moral domain.*J. Personal. Soc. Psychol.* 101, 2 (2011), 366.
[26] J. Habermas. 1995. Reconciliation through the public use of reason: Remarks on john rawls's political liberalism. *The Journal of Philosophy* 92, 3 (1995), 109–131.
[27] A. Hirschman. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, Harvard University Press.
[28] Jeroen van den Hoven, Dirk Helbing, Dino Pedreschi, Josep Domingo-Ferrer, Fosca Gianotti, and Markus Christen. 2012. FuturICT-The road towards ethical ICT. *The European Physical Journal Special Topics* 214, 1 (2012), 153–181.
[29] Nicole M. A. Huijts, Cees J. H. Midden, and Anneloes L. Meijnders. 2007. Social acceptance of carbon dioxide storage. *Energy Policy* 35, 5 (2007), 2780–2789.
[30] Dariusz Jemielniak. 2014. How to Fix Wikipedia's Bureaucracy Problem: Ignore All the Rules. Retrieved May 20, 2017 from http://www.slate.com/articles/technology/future_tense/2014/06/wikipedia_s_bureaucracy_problem_and_how_to_fix_it.html.
[31] Aniket Kittur and Robert E. Kraut. 2008. Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. ACM, 37–46.
[32] Mark Klein. 2012. Enabling large-scale deliberation using attention-mediation metrics. *Comput. Support. Coop. Work* 21, 4–5 (2012), 449–473.
[33] Christoph Carl Kling, Jérôme Kunegis, Heinrich Hartmann, Markus Strohmaier, and Steffen Staab. 2015. Voting behaviour and power in online democracy: A study of liquidfeedback in germany's pirate party. arXiv Preprint arXiv: 1503.07723 (2015).
[34] Bertram F. Malle. 2015. Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics and Information Technology*. 1–14.
[35] David C. Parkes and Lirong Xia. 2012. A complexity-of-strategic-behavior comparison between schulze's rule and ranked pairs. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI'12)*. American Association for Artificial Intelligence.
[36] J. Rawls. 1995. Political liberalism: Reply to habermas. *The Journal of Philosophy* 92, 3 (1995), 132–180.
[37] Lambèr Royakkers and Sjef Orbons. 2015. Design for values in the armed forces: Nonlethal weapons and military robots. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*. 613–638.
[38] Markus Schulze. 2011. A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Soc. Choice Welfare* 36, 2 (2011), 267–303.
[39] Amartya Kumar Sen. 1999. Democracy as a universal value. *J. Democr.* 10, 3 (1999), 3–17.
[40] I. Van de Poel. Forthcoming. A coherentist view on social acceptance and moral acceptability of technology. In *The Empirical Turn Revisited*, P. Vermaas, M. Franssen, and P. Froes (Eds.). Springer, The Netherlands.
[41] Ilse Verdiesen, Martijn Cligge, Jan Timmermans, Lennard Segers, Virginia Dignum, and Jeroen van den Hoven. 2016. MOOD: Massive open online deliberation platform a practical application. In *Proceedings of the 1st Workshop on Ethics in the Design of Intelligent Agents*. 6–11.
[42] Sheng Kung Michael Yi, Mark Steyvers, Michael D. Lee, and Matthew J. Dry. 2012. The wisdom of the crowd in combinatorial problems. *Cogn. Sci.* 36, 3 (2012), 452–470.

Received January 2017; revised December 2017; accepted January 2018